Revisiting Paulson's Theory of the Constructible Universe with Isar and Sledgehammer

Ioanna M. Dimitriou H. and Peter Koepke, University of Bonn, Germany

AITP 2016

Obergurgl, Austria, April 3-7, 2016

universitätbonn

Revisiting Paulson's Theory of the Constructible Universe - Natural proofs with Isabelle, Isar, Sledgehammer, and Naproche

Ioanna M. Dimitriou H. and Peter Koepke, University of Bonn, Germany

AITP 2016

Obergurgl, Austria, April 3-7, 2016

A vision: natural language mathematical proofs which are fully formal and proof checked

Example in SAD (Andrey Paskevich) with LaTeX sugar:

The *power set* of A is the set of subsets of A. Let $\mathcal{P}(A)$ denote the power set of A.

Theorem 1. (Cantor) *There is no surjection from A onto the power set of A.*

Proof. Assume F is a surjection from A onto $\mathcal{P}(A)$. Let

$$B = \{x \in A \mid x \notin F(x)\}.$$

$B \in \mathcal{P}(A)$. Take $a \in A$ such that $B = F(a)$.

$$a \in B \text{ iff } a \notin F(a) \text{ iff } a \notin B.$$

Contradiction. □

Naproche: Natural language proof checking

- combining formal mathematics with mathematical texts in natural language
- joint project with M. Cramer and B. Schröder
- NLP defining a controlled natural language and transforming input into FOL
- bridging proof gaps with strong ATPs like E or Vampire

A Naproche text

Axiom. *There is a set* $\emptyset$ *such that no* y *is in* $\emptyset$.

Axiom. *For every* x *it is not the case that* $x \in x$.

Define x to be *transitive* if and only if for all u, v, if $u \in v$ and $v \in x$ then $u \in x$. Define x to be an *ordinal* if and only if x is transitive and for all y, if $y \in x$ then y is transitive.

Theorem. For all x, y, if $x \in y$ and y is an ordinal then x is an ordinal.

Proof. Suppose $x \in y$ and y is an ordinal. Then for all v, if $v \in y$ then v is transitive. Hence x is transitive. Assume that $u \in x$. Then $u \in y$, i.e. u is transitive. Thus x is an ordinal. □

Theorem. (Burali-Forti) There is no x such that for all u, $u \in x$ iff u is an ordinal.

Proof. Assume for a contradiction that there is an x such that for all u, $u \in x$ iff u is an ordinal.

Lemma. x is an ordinal.

Proof. Let $u \in v$ and $v \in x$. Then v is an ordinal, i.e. u is an ordinal, i.e. $u \in x$. Thus x is transitive. Let $v \in x$. Then v is an ordinal, i.e. v is transitive. Thus x is an ordinal. Qed.

Then $x \in x$. Contradiction. □

"Revisiting" project

- Combining Naproche and Isabelle
- "Naturalizing" a comprehensive Isabelle formalization
- Larry Paulson's formalization of the constructible universe
- 1. Phase: rewriting proofs with Isar, using Sledgehammer
- 2. Phase: Interfacing Isabelle with Naproche

Overview

- Zermelo-Fraenkel set theory
- Axiomatic set theory
- Gödel's relative consistency of the Axiom of Choice
- Larry Paulson's formalization
- Formalizing axiomatic set theory
- Natural formalizations with Isar and Sledgehammer
- A HOL/FOL problem

Zermelo-Fraenkel set theory

1.1. ***Axiom of Extensionality.*** *If X and Y have the same elements, then $X = Y$.*

1.2. ***Axiom of Pairing.*** *For any a and b there exists a set $\{a, b\}$ that contains exactly a and b.*

1.3. ***Axiom Schema of Separation.*** *If P is a property (with parameter p), then for any X and p there exists a set $Y = \{u \in X : P(u, p)\}$ that contains all those $u \in X$ that have property P.*

1.4. ***Axiom of Union.*** *For any X there exists a set $Y = \bigcup X$, the union of all elements of X.*

1.5. ***Axiom of Power Set.*** *For any X there exists a set $Y = P(X)$, the set of all subsets of X.*

1.6. ***Axiom of Infinity.*** *There exists an infinite set.*

1.7. ***Axiom Schema of Replacement.*** *If a class F is a function, then for any X there exists a set $Y = F(X) = \{F(x) : x \in X\}$.*

1.8. ***Axiom of Regularity.*** *Every nonempty set has an $\in$-minimal element.*

The ZF axioms in first-order logic

a) $\exists x \forall y \neg y \in x$

b) $\forall x \forall y (\forall z (z \in x \leftrightarrow z \in y) \rightarrow x = y)$

c) $\forall x \forall y \exists z \forall w \, (u \in z \leftrightarrow u = x \vee u = y)$

d) $\forall x \exists y \forall z (z \in y \leftrightarrow \exists w (w \in x \wedge z \in w))$

e) $\forall x_1 ... \forall x_n \forall x \exists y \forall z \, (z \in y \leftrightarrow z \in x \wedge \varphi(z, x_1, ..., x_n))$

f) $\forall x \exists y \forall z (z \in y \leftrightarrow \forall w (w \in z \rightarrow w \in x))$

g) $\forall x_1 ... \forall x_n (\forall x \forall y \forall y' ((\varphi(x, y, x_1, ..., x_n) \wedge \varphi(x, y', x_1, ..., x_n)) \rightarrow y = y') \rightarrow \forall u \exists v \forall y \, (y \in v \leftrightarrow \exists x (x \in u \wedge \varphi(x, y, x_1, ..., x_n))))$

h) $\exists x (\exists y \, (y \in x \wedge \forall z \neg z \in y) \wedge \forall y (y \in x \rightarrow \exists z (z \in x \wedge \forall w (w \in z \leftrightarrow w \in y \vee w = y))))$

i) $\forall x_1 ... \forall x_n (\exists x \varphi(x, x_1, ..., x_n) \rightarrow \exists x (\varphi(x, x_1, ..., x_n) \wedge \forall x' (x' \in x \rightarrow \neg \varphi(x', x_1, ..., x_n))))$

ZF - a foundation for mathematics

K. Gödel, *Über formal unentscheidbare Sätze der Principia mathematica ...* (1931):

The development of mathematics towards greater precision has led, as is well known, to the formalization of large tracts of it, so that one can prove any theorem using nothing but a few mechanical rules. The most comprehensive formal systems that have been set up hitherto are the system of *Principia mathematica* (PM) on the one hand and the Zermelo-Fraenkel axiom system of set theory. These two systems are so comprehensive that in them all methods of proof today used in mathematics are formalized, that is, reduced to a few axioms and rules of inference.

ZF - a foundation for mathematics

- ZF or ZF with the Axiom of Choice (ZFC) covers all (or 99%) of mathematics

- the formalizability of mathematics in ZF(C) is a basis for the programme of *Formal Mathematics*

- it is difficult to come up with notions that are not covered by ZF(C)

- is every mathematical statement decided True or False by ZF(C)?

A "logical" incompleteness of ZF

Gödel incompleteness theorem:

If ZF is a consistent theory then there is a (number theoretic) statement φ which codes the unprovability of itself in ZF, such that ZF proves neither φ nor $\neg\varphi$.

A mathematical incompleteness of ZF

The Axiom of Choice (AC): every set x possesses a well-order $<$ (so that induction over all elements of x along $<$ is possible)

Paul J. Cohen (1963):

If ZF is consistent then ZF proves neither AC nor $\neg$AC.

Gödel had already proved (1940):

If ZF is consistent then ZF does not prove $\neg$AC (The *relative consistency* of the axiom of choice; Con(ZF) $\rightarrow$ Con(ZF+AC))

Cohen's result marks the start of modern *axiomatic set theory*. Thousands of independence results have been proved using Gödel's method of the *constructible universe* and generalizations, and using Cohen's *forcing method*.

Paulson's formalization of Gödel's relative consistency result

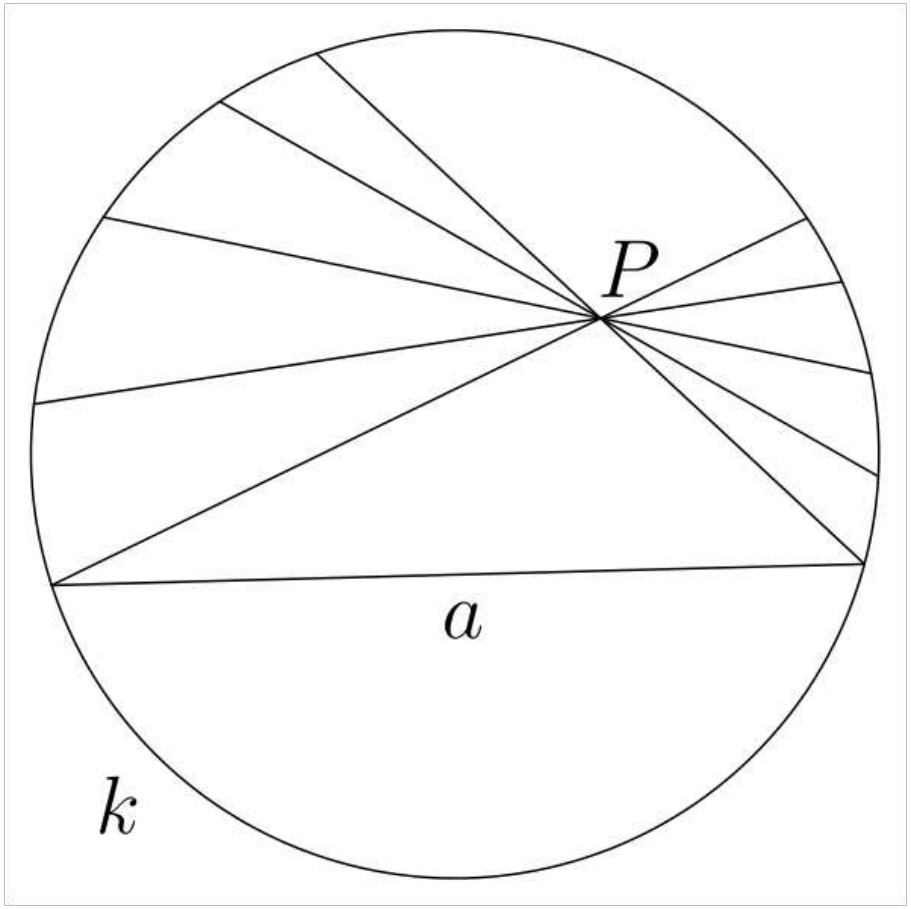
Formalizing modern set theory means formalizing relative consistency results

In 2003, Paulson formalized the relative consistency of AC, using Gödel's constructible universe L:

```
theorem "∀x[L]. ∃r. wellordered(L,x,r)"
proof
  fix x
  assume "L(x)"
  then obtain r where "well_ord(x,r)"
    by (blast dest: L_implies_AC)
  thus "∃r. wellordered(L,x,r)"
    by (blast intro: well_ord_imp_relativized)
qed
```

Inner models

Beltrami-Klein model for hyperbolic geometry is an "inner model" of the euclidean plane



The constructible universe

The inner model of constructible sets (from Paulson, 2003)

The consistency of AC, mechanized

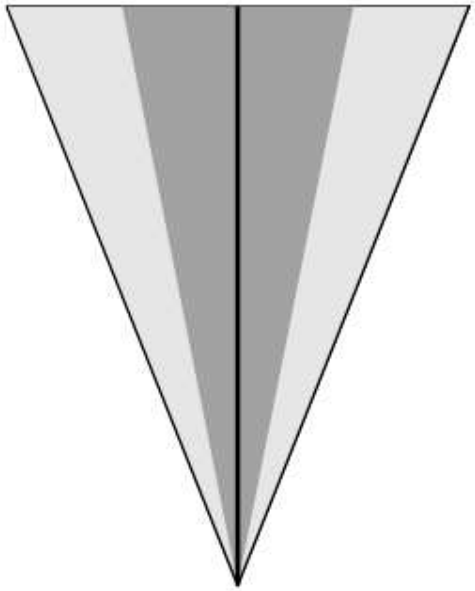


Figure 1: The constructible universe, **L**.

Gödel's relative consistency proof

Paulson, 2003:

> Gödel's proof involves four main tasks:
>
> 1. defining the class **L** within ZF;
> 2. proving that **L** satisfies the ZF axioms;
> 3. proving that **L** satisfies **V**=**L**;
> 4. proving that **V**=**L** implies the axiom of choice.

Gödel's relative consistency proof

Paulson, 2003:

> Gödel's proof involves four main tasks:
>
> 1. defining the class **L** within ZF;
> 2. proving that **L** satisfies the ZF axioms;
> 3. proving that **L** satisfies **V**=**L**;
> 4. proving that **V**=**L** implies the axiom of choice.

Gödel's relative consistency proof

Paulson, 2003:

> Gödel's proof involves four main tasks:
>
> 1. defining the class **L** within ZF;
> 2. proving that **L** satisfies the ZF axioms;
> 3. proving that **L** satisfies **V**=**L**;
> 4. proving that **V**=**L** implies the axiom of choice.

Gödel's relative consistency proof

Paulson, 2003:

> Gödel's proof involves four main tasks:
>
> 1. defining the class **L** within ZF;
> 2. proving that **L** satisfies the ZF axioms;
> 3. proving that **L** satisfies **V**=**L**;
> 4. proving that **V**=**L** implies the axiom of choice.

Formal development of set theory

Elliott Mendelson, *Introduction to Mathematical Logic*

Proposition 4.9 (Transfinite Induction)

$$\vdash (\forall\beta)\big[(\forall\alpha)(\alpha\in\beta\Rightarrow\alpha\in X)\Rightarrow\beta\in X\big]\Rightarrow On\subseteq X$$

(If, for every β, whenever all ordinals less than β are in X, β must also be in X, then all ordinals are in X.)

Proof

Assume $(\forall\beta)\ [(\forall\alpha)(\alpha\in\beta\Rightarrow\alpha\in X)\Rightarrow\beta\in X]$. Assume there is an ordinal in $On-X$. Then, since On is well-ordered by E, there is a least ordinal β in $On-X$. Hence, all ordinals less than β are in X. So, by hypothesis, β is in X, which is a contradiction.

Formalizing with Isabelle, Isar, Sledgehammer

```
lemma Transfinite_Induction: "(∀β∈On.(∀α∈On.(α∈β⟶α∈X)⟶β∈X))⟶On⊆X"
proof (rule impI, rule ccontr)
  assume premise: "∀β∈On.(∀α∈On. ((α∈β⟶α∈X)⟶β∈X)))"
  assume contra:  "¬(On⊆X)"
  hence *: "∃γ∈On. γ∈(On\X)" using Ex4_9_c exists_ordinal_def set_subclassI
by fastforce
  hence "(On\X) has a least element with respect to ℰ" using Prop4_8_f
  proof -
    have "(On\X)≠∅" using * NBG_Set.empty_set exists_ordinal_def by auto
    moreover have "(On\X)⊆On" using B2 set_subclassI by blast
     thus ?thesis using Prop4_8_f unfolding Well_ord_of_def using calculation
by blast  qed
   then obtain β where **: "β is the least in (On\X) with respect to ℰ" by
auto
   hence "∀γ∈On. (γ<β⟶γ∈X)" using premise Ex4_31_a NBG_Set.empty_set
forall_ordinals_def by auto
   thus False using premise unfolding less_on_ordinals_def using Ex4_31_a
Ex4_9_c    NBG_Set.empty_set    Rep_Set_inverse    **    forall_ordinals_def
notin_inter_mono by auto
qed
```

Isabelle, Isar, tactics obtained by Sledgehammer hidden

```
lemma Transfinite_Induction: "(∀β∈On.(∀α∈On.(α∈β⟶α∈X)⟶β∈X))⟶On⊆X"
proof (rule impI, rule ccontr)
  assume premise: "∀β∈On.(∀α∈On. ((α∈β⟶α∈X)⟶β∈X)))"
  assume contra:  "¬(On⊆X)"
  hence *: "∃γ∈On. γ∈(On\X)"
  hence "(On\X) has a least element with respect to ℰ"
  proof -
    have "(On\X)≠∅"
    moreover have "(On\X)⊆On"
    thus ?thesis
    qed
  then obtain β where **: "β is the least in (On\X) with respect to ℰ"
  hence "∀γ∈On. (γ<β⟶γ∈X)"
  thus False
qed
```

Isabelle, Isar, Sledgehammer, NLP

Lemma (Transfinite Induction). Assume $\forall\beta \in \mathrm{On}.(\forall\alpha \in \mathrm{On}.(\alpha \in \beta \rightarrow \alpha \in X) \rightarrow \beta \in X)$. Then $\mathrm{On} \subseteq X$.

Proof (using impI, ccontr). Assume $\mathrm{On} \nsubseteq X$. Hence $\exists\gamma \in \mathrm{On}.\gamma \in (\mathrm{On} \setminus X)$.
Claim. $(\mathrm{On} \setminus X)$ has a least element with respect to $\mathcal{E}$.
Proof. $(\mathrm{On} \setminus X) \neq \emptyset$. $(\mathrm{On} \setminus X) \subseteq \mathrm{On}$. This implies the thesis. qed(Claim)

Then take some β such that β is the least element of $(\mathrm{On} \setminus X)$ with respect to $\mathcal{E}$. Hence $\forall\gamma \in \mathrm{On}.(\gamma < \beta \rightarrow \gamma \in X)$. Contradiction. $\square$

Isabelle, Isar, Sledgehammer, Naproche

- Isabelle: powerful proof assistant with comprehensive libraries

- Isar: language for structured proofs

- Sledgehammer: bridging simple proof steps

- Naproche-style natural language processing

- HOL with quantifications over formulas is beneficial for the FOL theory ZF: ZF uses schemas of axioms, definition, lemmas; Con(ZF) $\rightarrow$ Con(ZF+AC) is a HOL statement

Cantor's theorem in Isabelle and SAD

```
lemma cantor: "∃S ∈ Pow(A). ∀x∈A. b(x) ≠ S"
by (best elim!: equalityCE del: ReplaceI RepFun_eqI)
end
```

Theorem 2. (Cantor) *There is no surjection from A onto the power set of A.*

Proof. Assume F is a surjection from A onto $\mathcal{P}(A)$. Let

$$B = \{x \in A \mid x \notin F(x)\}.$$

$B \in \mathcal{P}(A)$. Take $a \in A$ such that $B = F(a)$.

$$a \in B \text{ iff } a \notin F(a) \text{ iff } a \notin B.$$

Contradiction. □

Problems

- Sledgehammer only available for Isabelle-HOL, not for FOL

- ZF is FOL

- ZF formalized in HOL a stronger theory than first-order ZF

- S. Agerholm and M.J.C. Gordon. *Experiments with ZF Set Theory in HOL and Isabelle* (1995)

AC in HOL

Isabelle HOL includes Hilbert's choice operator:

```
subsection ‹Hilbert's epsilon›

axiomatization Eps :: "('a => bool) => 'a" where
  someI: "P x ==> P (Eps P)"
```

This implies various versions of Choice

```
subsection‹Axiom of Choice, Proved Using the Description Operator›

lemma choice: "∀x. ∃y. Q x y ==> ∃f. ∀x. Q x (f x)"
by (fast elim: someI)

lemma bchoice: "∀x∈S. ∃y. Q x y ==> ∃f. ∀x∈S. Q x (f x)"
by (fast elim: someI)

lemma choice_iff: "(∀x. ∃y. Q x y) ⟷ (∃f. ∀x. Q x (f x))"
```

Set theory in HOL

ZF*: A natural axiomatization of set theory in HOL

Extensionality: $\forall s\ t.\ (s = t) = (\forall x.\ x \in s = x \in t)$

Empty set: $\exists s.\ \forall x.\ \neg(x \in s)$

Union: $\forall s.\ \exists t.\ \forall x.\ x \in t = (\exists u.\ x \in u \wedge u \in s)$

Power sets: $\forall s.\ \exists t.\ \forall x.\ x \in t = x \subseteq s$

Separation: $\forall p\ s.\ \exists t.\ \forall x.\ x \in t = x \in s \wedge p\ x$

Replacement: $\forall f\ s.\ \exists t.\ \forall y.\ y \in t = \exists x.\ x \in s \wedge (y = f\ x)$

Foundation: $\forall s.\ \neg(s = \emptyset) \Rightarrow \exists x.\ x \in s \wedge (x \cap s = \emptyset)$

Infinity: $\exists s.\ \emptyset \in s \wedge \forall x.\ x \in s \Rightarrow (x \cup \{x\}) \in s$

On the strength of ZF*

- ZF* $\vdash$ ZF

- ZF* $\vdash$ AC

- Gödel's model L is defined by iterating FOL-definability

- L is not a model of ZF* but of ZF

- should we define a HOL-based L* ?

- ...

ZF in Isabelle-HOL

- Proof-theoretic results need exact axiomatic strength

- to prove Con(ZF) $\rightarrow$ Con(ZF+AC) one has to start the Gödel construction exactly from the ZF axioms

A workaround:

- use the von Neumann - Bernays - Gödel class theory NBG instead of ZF; NBG is a conservative extension of ZF

- NBG is finitely axiomatizable in HOL

- ...

Summary

- Sledgehammer introduces valuable AI into Isabelle and helps to write concise Isar proofs

- Sledgehammer is often able to connect "natural" proof steps in textbook proofs, so that "natural" Isar proofs can be built with Isabelle and Sledgehammer

- Problems: Sledgehammer requires HOL, jeopardizing the proof-theoretic applicability of Isabelle proofs, etc. etc.)

Thank you, and enjoy the snow!